DOCUMENT RESUME

ED 178 168 PS 010 927

AUTHOR King, Daniel W.; King, Lynda A.
TITLE [The Developmental Generalizability of Scales Designed to Evaluate Parent-Child Interactions: Technical Reports on the Nursing Child Assessment Teaching and Feeding Scales.]
INSTITUTION Washington Univ., Seattle.
PUB DATE Mar 78
NOTE 44p.; Paper presented at the Biennial Conference of the Society for Research in Child Development (San Francisco, California, March 15-18, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Food; Infant Behavior; Longitudinal Studies; *Measurement Instruments; Measurement Techniques; *Nonverbal Communication; *Parent Child Relationship; *Rating Scales; *Reliability; Research Problems; Sampling; *Verbal Communication
IDENTIFIERS Nursing Child Assessment Feeding Scales; Nursing Child Assessment Teaching Scales

ABSTRACT

This paper presents results of a generalizability assessment of two sets of scales designed to evaluate the quality of verbal and nonverbal interactions between parent and child. The Nursing Child Assessment Feeding Scale (NCAFS) and Nursing Child Assessment Teaching Scales (NCATS) were developed as part of a longitudinal study aimed at determining early indicators of child development. Designed for use by trained observers, both sets of scales are Likert-type and both provide individual scale scores plus cluster scores reflecting the measure of constructs believed important to parent-child interactions. Both scales were administered to children and parents who were members of a health care cooperative in Seattle. The Feeding Scales were given four times during the child's first year and the Teaching Scales four times during the child's first two years. Results of generalizability analyses suggested further research with the instrument with more heterogeneous samples. Also, findings indicated that increasing the number of administrations across time could best decrease measurement error. Detailed technical reports on the reliability of the scales are included. (Author/RH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED178168

THE DEVELOPMENTAL GENERALIZABILITY OF SCALES DESIGNED TO EVALUATE PARENT-CHILD INTERACTIONS: TECHNICAL REPORTS ON THE NURSING CHILD ASSESSMENT TEACHING AND FEEDING SCALES. (Title Supplied)

The Developmental Generalizability of Scales Designed to Evaluate Parent-Child Interactions

Daniel W. King
Central Michigan University

Lynda A. King
University of Washington

Paper presented as part of a symposium (Generalizability Studies of Mother and Infant Behavior) at the 1979 Biennial Conference of the Society for Research in Child Development, March 15-18, 1979, San Francisco.

PS 010927

This paper is based on data obtained when both authors were associated with the Nursing Child Assessment Follow-up Project, University of Washington. Copies may be obtained by writing:

Dr. Sandra Eyres
NCAF
Maternal-Child Nursing
School of Nursing
University of Washington
Seattle, Washington 98195

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Daniel W. King

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Developmental Generalizability of Scales Designed to Evaluate Parent-Child Interactions

We would like to present the results of a generalizability assessment of two sets of scales designed to evaluate the quality of the interactions between parent and child. The scales were developed in the context of the Nursing Child Assessment longitudinal studies at the School of Nursing, University of Washington, Seattle, Washington. The Nursing Child Assessment Project (NCAP) and its sequel, the Nursing Child Assessment Follow-up (NCAF), have concentrated on determining early indicators of later child development outcomes. Somewhat less than 200 parents and their first-born children have been followed now up to the child's fifth year with a series of in-depth assessments at designated time points. Many of these assessments have been conducted in the home. All parent-child participants are members of a health maintenance cooperative in Seattle.

Instrumentation has been a concern throughout the studies, and a variety of instruments have been used. The two instruments on which we are reporting today were developed by Dr. Kathryn Barnard and her associates to measure what they believed to be an essential area upon which to focus in predicting developmental outcomes: the nature of the verbal and nonverbal communication between parent and child.

The original Nursing Child Assessment Teaching Scales (NCATS) are a series of 24 five-point Likert-type scales on which trained observers give behavioral ratings during a teaching situation, that is, when the parent is teaching the child selected tasks. The instrument has been administered twice at

each assessment point--first, when the parent is asked to teach a task considered at age level for the child, and second, when the parent is asked to teach a task considered above age level for the child.

One- or two-word identifiers for the 24 teaching scales are shown on page one of your handout. The first nine scales assess the child's behavior and/or response to the parent, and the remaining scales assess the parent's behavior and/or response to the child. To provide you with a better understanding of the NCATS items, I will read an example from the instrument, Scale #9, "Alertness," a child-assessment item:

> Alertness as conveyed through facial expression. This is an "animated facial expression" scale in which consideration of physical activity should be excluded from the ratings. Ratings should be based on instances of maximum animation.
>
> [0. Impossible to rate because of prolonged crying.]
>
> 1. Dull; unresponsive to stimuli. Minimal focusing on either mother, task materials, or features of room.
>
> 2. Somewhat alert.
>
> 3. Infant seems focused but without enthusiasm.
>
> 4. Alert. Infant focused with some enthusiasm.
>
> 5. Infant alert. Focused with intense enthusiasm.

As page one of your handout also shows, in addition to individual scales, a set of five teaching clusters was also developed, each comprised of items believed to represent single important constructs related to parent-child interactions. Cluster or composite scores are determined as the sum of item scores within each cluster, and these composite scores have been used quite extensively in the research projects.

The second instrument with which we were concerned is the Nursing Child Assessment Feeding Scales (NCAFS). The names of the 21 feeding scales and

the feeding clusters are listed on page two of the handout. Like the NCATS, the NCAFS are used by trained observers to rate both parent and child behaviors, but in this case, in the situation where the child is being fed by the parent. The scales have been used to observe breast-feeding, bottle-feeding, as well as the feeding of solid foods.

The feeding scales are seven-point Likert-type scales. To give you an idea of their format and content, I will read you the scale labeled "Kinesthetic," Scale #3 on the handout, a parent-assessmert item:

[0. Not applicable - e.g., infant positioned so as to preclude appropriate opportunities for kinesthetic stiumlation.]

1. No kinesthetic stimulation.
2. Minimal amount of kinesthetic stimulation as appropriate for age.
3. Some kinesthetic stimulation as appropriate for age.
4. Position movement for burping, rocking, and position change; motion variable; intensity age-appropriate.
5. Much kinesthetic stimulation as appropriate for age.
6. Excessive amount of kinesthetic stimulation provided as appropriate for age.
7. Kinesthetic stimulation continuous, obligatory, nonpurposeful in regard to infant.

As the definitions of the points on this seven-point "Kinesthetic" scale reflect, the NCAFS are designed such that scores of 1 and 7, the extremes, represent what we regard as maladaptive behaviors, whereas the mid-point score of 4 represents the most appropriate or adaptive behaviors. Thus, when we desired to obtain cluster scores for the feeding scales, it was necessary to "fold" the scales at the mid-point value of 4 before summative scores were calculated. A score of 7 became 1, a score of 6 became 2, and a score of 5 became 3. Composite scores were then obtainable to measure constructs

identified as child's "Sensitivity to Cues," parent's "Fostering of Growth," and parent's "Alleviation of Distress."

Essentially, the measures of parent-child interactions available from the two instruments were of four forms: (a) individual teaching scales, (b) individual feeding scale, (c) teaching clusters, and (d) feeding clusters. In turn, a generalizability assessment of the precision of measurement required the use of four separate generalizability designs and analyses, which we will now review.

Very briefly, generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) is a multidimensional or "multifaceted" approach to the estimate of reliability (precision or accuracy) of a given measurement instrument. Instead of examining, for example, test-retest reliability or internal consistency reliability separately, the generalizability approach treats a number of dimensions simultaneously and allows one to make statements about the precision of a measure across more than one dimension at a time. The information from a generalizability study also permits one to make recommendations about future uses of the instrument under consideration, to capitalize on its strengths and weaknesses. In a generalizability study, the dimensions over which you wish to generalize are called "facets," and the data are arranged and analyzed as in an analysis of variance design. The emphasis, however, is on components of variance rather than on expected mean squares and tests of significance.

Pages three, four, and five of your handout present some basic information on the designs used to assess the NCATS and NCAFS. With regard to the teaching scales, the parent-child dyads were observed on five separate occasions (when the child was 1, 4, 8, 12, and 24 months of age), and they were

observed teaching both an easy (at age level) and a difficult (above age level) task. Thus, the design had two facets, age or occasion and task difficulty. These were fully crossed with one another and with subjects or dyads, and subjects and facets were treated as random. The estimate of universe score variance and components of observed score variance are shown on the handout. Because the instrument is essentially designed for clinical use, the components comprising the estimate of measurement error are those shown.

For the teaching clusters or composites, another facet, items, representing all scales within a cluster, was added. This facet was fully crossed with the other two facets, forming a three-facet, fully crossed design having the components listed on pages three and four.

For the individual feeding scales, the parent-child dyads were observed on four occasions (1, 4, 8, and 12 months). At 1 and 4 months of age, the children were given milk, and at 8 and 12 months of age, they were observed being fed solid foods by the parents. We felt that type of food (liquid or solid) might be an important variable, so it was included as a facet. Furthermore, because we believed that the levels "liquid" and "solid" exhausted the population or domain of food types, we treated it as a fixed facet in the analyses. Thus, the design for the individual NCAFS had age of child nested within food type and these crossed with subjects or dyads.

Because of incomplete data in analyzing the NCAFS clusters, we could only use the 1 and 4 month levels of the age of child facet. Items were fully crossed with ages and subjects, and subjects and facets were considered random. Your handout shows the component(s) comprising the estimate of universe score variance, observed score variance, and measurement error.

On page six, you will find summary data for the various generalizability analyses. The rows of the table are components of variance and the columns are the scales or clusters under consideration. The table presents ranges of values. The most striking result of our analyses was low values for the subject component of variance, that representing the estimate of universe score variance. This component represents the ability of the scales (or clusters) to identify individual differences. Because of the homogeneity of the sample on which these data are based (members of a cooperative health-maintenance organization, with documented similarities on many important organismic variables), it is not clear at this point whether the low values for the subject component reflect this homogeneity or insensitivity of the scales to individual differences in parent-child dyadic interactions. Further research with the NCATS and NCAFS on more demographically heterogeneous samples is required to resolve this issue.

The large values of the age component, both alone and in interaction with subjects and other facets, suggest that an accurate estimate of a dyad's universe score could best be obtained by finding an average across many administrations.

Finally, the results of the analyses seem to suggest that the inclusion of a task difficulty facet for the NCATS individual items and clusters and a food type facet for the NCAFS individual items add relatively little to the precision of the respective instruments and the information that can be generated from them.

COMPLETE REPORTS OF THE GENERALIZABILITY ANALYSES OF THE NCATS AND NCAFS MAY BE OBTAINED BY WRITING DR. SANDRA EYRES (address on title sheet) AND REQUESTING:

Technical Report: Generalizability Study of the Nursing Child Assessment Teaching Scales

Technical Report: Generalizability Study of the Nursing Child Assessment Feeding Scales

References

Cronbach, L. J., Gleser, G., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley & Sons, 1972.

Teaching Scales and Clusters

Scale Number	Scale Name	Cluster Membership
1	Initial state	
2	Displeasure	
3	Attention	Cluster 3-6-7-9: Readiness to Learn
4	Vocalization	
5	Success	
6	Intensity	Cluster 3-6-7-9: Readiness to Learn
7	Duration	Cluster 3-6-7-9: Readiness to Learn
8	Activity	
9	Alertness	Cluster 3-6-7-9: Readiness to Learn
10	Pos. Feedback	Cluster 10-23: Positive Message
11	Neg. Feedback	Cluster 11-24: Negative Messages
12	Modeling	
13	Materials	Cluster 13-14-18-21: Facilitation
14	Positioning	Cluster 13-14-18-21: Facilitation
15	Directiveness	Cluster 15-16-17: Techniques
16	Forcing	Cluster 15-16-17: Techniques
17	Guidance	Cluster 15-16-17: Techniques
18	Timing	Cluster 13-14-18-21: Facilitation
19	Exploratory	
20	Comfort	
21	Sensitivity	Cluster 13-14-18-21: Facilitation
22	Style	
23	Affection	Cluster 10-23: Positive Messages
24	Disapproval	Cluster 11-24: Negative Messages

Feeding Scales and Clusters

Scale Number	Scale Name	Cluster Membership
1	Positioning	Cluster 1-2: Fostering of Growth
2	Focus	Cluster 1-2: Fostering of Growth
3	Kinesthetic	Cluster 3-4-5-6-7: Sensitivity to Cues
4	Visual (Parent)	Cluster 3-4-5-6-7: Sensitivity to Cues
5	Tactile	Cluster 3-4-5-6-7: Sensitivity to Cues
6	Verbal I	Cluster 3-4-5-6-7: Sensitivity to Cues
7	Verbal II	Cluster 3-4-5-6-7: Sensitivity to Cues
8	Mood (Parent)	
9	Tension	
10	Response I	Cluster 10-11: Alleviation of Distress
11	Response II	Cluster 10-11: Alleviation of Distress
12	Initial State	
13	Predominant State	
14	State Changes	
15	Motor Activity	
16	Attentiveness	
17	Vocalization	
18	Mood (Child)	
19	Tone	
20	Visual (Child)	
21	Control	

GENERALIZABILITY DESIGNS: TEACHING SCALES

Individual Scales

2-facet, fully crossed design: s x d x a;

where: s = subjects

d = difficulty of task (levels = easy, difficult)

a = age of child (levels = 1, 4, 8, 12, 24 months);

subjects and facets are random.

Estimate of universe score variance: $\sigma^2(s)$

Components of observed score variance: $\sigma^2(s)$
$\sigma^2(sd)$
$\sigma^2(sa)$
$\sigma^2(sda,e)$

Components of error SQRT $\sigma^2(\Delta)$: $\sigma^2(d)$
$\sigma^2(a)$
$\sigma^2(sd)$
$\sigma^2(sa)$
$\sigma^2(sda,e)$

Clusters

3-facet, fully crossed design: s x d x a x i;

where: s = subjects

d = difficulty of task (levels = easy, difficult)

a = age of child (levels = 1, 4, 8, 12, 24 months)

i = items (levels = the scales in the cluster);

subjects and facets are random.

Estimate of universe score variance: $\sigma^2(s)$

Components of observed score variance: $\sigma^2(s)$
$\sigma^2(si)$
$\sigma^2(sd)$
$\sigma^2(sa)$
$\sigma^2(sid)$
$\sigma^2(sia)$
$\sigma^2(sda)$
$\sigma^2(sida,e)$

Components of error SQRT $\sigma^2(\Delta)$:

$\sigma^2(i)$
$\sigma^2(d)$
$\sigma^2(a)$
$\sigma^2(si)$
$\sigma^2(sd)$
$\sigma^2(sa)$
$\sigma^2(sid)$
$\sigma^2(sia)$
$\sigma^2(sda)$
$\sigma^2(sida,e)$

GENERALIZABILITY DESIGNS: FEEDING SCALES

Individual Scales

2-facet, split plot design: s x a(t);

where: s = subjects

a = age of child (levels = 1, 4, 8, 12 months)

t = food type (levels = milk, solid);

subjects and age of child are random, food type is fixed.

Estimate of universe score variance: $\sigma^2(s)$

Components of observed score variance: $\sigma^2(s)$
$\sigma^2(st)$
$\sigma^2(sta,sa,e)$

Components of error SQRT $\sigma^2(\Delta)$:

$\sigma^2(t)$
$\sigma^2(st)$
$\sigma^2(a,ta)$
$\sigma^2(sta, sa,e)$

Clusters

2-facet, fully crossed design: s x a x i;

where: s = subjects

a = age of child (levels = 1, 4 months)

i = items (levels = the scales in the cluster);

subjects and facets are random.

Estimate of universe score variance: $\sigma^2(s)$

Components of observed score variance: $\sigma^2(s)$
$\sigma^2(si)$
$\sigma^2(sa)$
$\sigma^2(sia,e)$

Components of error SQRT $\sigma^2(\Delta)$:
$\sigma^2(i)$
$\sigma^2(a)$
$\sigma^2(si)$
$\sigma^2(sa)$
$\sigma^2(sia,e)$

RANGE OF VALUES FOR VARIANCE COMPONENTS: 6

Variance Components	The 24 Teaching Scales	The 5 Teaching Clusters	The 21 Feeding Scales	The 3 Feeding Clusters
s	.000-.129	.031-.091	.000-.442	.000-.181
i		.000-1.015		.000-.063
d	.000-.183	.000-.042		
t			.000-.479	
a	.000-.500	.000-.150		.000-.005
si		.008-.026		.038-.162
sd	.000-.039	.000-.009		
st			.000-.233	
sa	.065-.400	.032-.176		.078-.580
id		.000-.031		
ia		.004-.074		.000-.019
da	.000-.261	.000-.110		
a,ta			.000-.192	
sid		.000-.011		
sia		.059-.163		
ida		.003-.049		
sda		.069-.292		
sia,e				.517-.730
sda,e	.177-.844			
sida,e		.214-.359		
sta,sa,e			.226-2.790	

Technical Report: Generalizability Study of the Nursing Child Assessment Teaching Scales

Daniel W. King and Lynda A. King

Nursing Child Assessment Follow-up Project

Maternal and Child Nursing

University of Washington

Seattle, Washington

November 1978

Daniel King is currently Assistant Professor, Department of Psychology, Central Michigan University, Mt. Pleasant, Michigan.

Technical Report: Generalizability Study of the Nursing Child Assessment Teaching Scales

The Nursing Child Assessment Teaching Scales (NCATS) consist of 24 separate 5-point Likert-type scales designed to evaluate the quality of parent-child interaction during a teaching situation. An observer makes behavioral ratings of both parent and child first when the parent teaches a task considered at age level for the child, and second when the parent teaches a task considered above age level for the child. Scores for each task are obtainable on each scale individually and on five scale clusters or composites of particular individual scales. Table 1 presents a list of the 24 scales and, where appropriate, the scales which comprise the five clusters. A detailed description of each scale and cluster and the theoretical rationale for its development may be found elsewhere (Barnard, 1978). For the purposes of this report, the scales will be identified by scale number as shown on Table 1, and the clusters will be identified by their component scale numbers.

The present report provides data on the precision or reliability of the NCATS from the perspective of generalizability theory as proposed by Cronbach, Gleser, Nanda, and Rajaratnam (1972). Briefly, a generalizability study allows for a multivariate or "multi-faceted" approach to reliability. That is, rather than looking at different types of reliability (test-retest, alternate forms) as independent or unrelated estimates of precision, the generalizability approach allows one to simultaneously examine the contributions of predetermined factors or "facets" to observed score variance using an analysis of variance paradigm. Such information enables those interested in using an

instrument to clarify specific sources of variance and thus to develop designs which are more efficient in enhancing the accuracy of the instrument. In addition, a multifaceted study enables one to evaluate interactions among facets; knowledge of these interaction effects, which are not obtainable using traditional procedures, improves one's understanding of the instrument.

Data for this generalizability analysis were generated as a part of a longitudinal study to identify significant child development variables and to propose procedures by which nurses can measure these variables. The subjects in the longitudinal study were parents and children who held membership in a health-care cooperative in Seattle, Washington. Generalizability data were taken from measures obtained at five time points in the child's first two years: at ages 1 month, 4 months, 8 months, 12 months, and 24 months. The NCATS were administered as a part of a battery of instruments by trained professionals during assessment sessions conducted in the homes of participants.

Individual Scales

A two-facet fully-crossed design was used to evaluate the precision of each individual scale. The design is represented by $\underline{s} \times \underline{d} \times \underline{a}$, where $\underline{s}$ refers to subjects, $\underline{d}$ refers to a facet identified as difficulty of task (easy, difficult), and $\underline{a}$ refers to a facet identified as age of child (1 month, 4 months, 8 months, 12 months, 24 months). The numbers of subjects involved in the analyses differ from scale to scale (see Table 2). For any single scale analysis, complete ratings of subjects at all ages for both easy and difficult tasks were required, and subjects with incomplete data were necessarily eliminated from an analysis. In the $\underline{s} \times \underline{d} \times \underline{a}$ design, subjects ($\underline{s}$), difficulty of task ($\underline{d}$), and age of child ($\underline{a}$) were all treated as random.

Table 2 presents components of variance, errors of measurement, and generalizability coefficients for each scale. Consistent with the notation used by Cronbach et al. (1972), parenthetical lower-case Roman letters indicate sources of variance for components. Errors of measurement are identified by SQRT $\sigma^2(\varepsilon)$, SQRT E $\sigma^2(\delta)$, and SQRT $\sigma^2(\Delta)$, and the generalizability coefficient $E\rho^2$ is interpreted as the ratio of universe score variance to observed score variance.

Perhaps the most striking result of the analyses is the relatively low values obtained for the subject (s) component of variance. With a 5-point scale, a desirable value for $\sigma^2(s)$ is somewhere between 1.000 and 1.500, which would indicate the full range of the scale is used. For these analyses, the range of $\sigma^2(s)$ from .000 to .129 implies that most of the ratings were within two points or less of each other. This finding suggests one or both of the following situations: (a) The scales are either insensitive to individual differences in parent-child interactions, and/or (b) the sample from which these data are derived is extremely homogeneous. One might argue that sample homogeneity provides a weak rationale for the inability to identify individual differences. This issue remains for future empirical examination.

Because the NCATS are designed primarily for use in individual assessment, SQRT $\sigma^2(\Delta)$ is the appropriate estimate of measurement error (Cronbach et al., 1972). On first examination, the values of SQRT $\sigma^2(\Delta)$ appear encouraging given the 5-point range of the scales. One should, however, exercise caution in interpretation because of the limited sizes of the subject components of variance.

Relative to the other components, variability due to task difficulty [reflected in the sizes of the $\sigma^2(d)$, $\sigma^2(da)$, and $\sigma^2(sd)$ components] is small.

One might conclude that adding conditions by increasing the number of task difficulty levels will have a negligible influence on both the generalizability coefficient and measurement error. Furthermore, the distinction between easy and difficult tasks as presently used seems to add relatively little to the precision of the scales and the information which can be generated from them.

Finally, the a facet (age of child) contributes an appreciable amount of variability either in interaction with s [σ^2(sa)] or as a part of the residual [σ^2(sda,e)]. Since both of these components are included in estimates of error SQRT $\sigma^2(\Delta)$, more precise estimates of universe scores would be expected if the number of observations over the two-year period were increased. Table 4 gives generalizability coefficients and estimates of error when N_a = 10, rather than 5, observations of the parent-child interaction are obtained. As expected, with the exception of Scale 12, error estimated by SQRT $\sigma^2(\Delta)$ decreases while values of $E\rho^2$ increase.

Clusters

To evaluate the teaching clusters, a three-facet fully-crossed design was used: s x d x a x i, where s again refers to subjects, d again refers to task difficulty (easy, difficult), a again refers to age of child (1 month, 4 months, 8 months, 12 months, 24 months), and i (items) represents a facet with conditions being the several scales comprising a cluster. Subjects (s) and all facets (d, a, and i) were considered random. Table 5 presents the number of subjects involved in each analysis and the number of conditions of facet i in each analysis.

Table 6 presents the results of the analyses. As with the analyses of the individual scales, variability due to subjects [σ^2(s)] is relatively

small. Again, the values of SQRT $\sigma^2(\Delta)$ must be evaluated in terms of this small subject variability. Similar to the residual component in the previous analyses, the component σ^2(sda) emerges as a strong contributor to variance. Depending upon the age of child at instrument administration, subjects responded differently to the different tasks. Likewise, the relatively large σ^2(sia) for Cluster 15-16-17, Cluster 10-23, and Cluster 11-24 indicates that subjects responded differently to the different items depending upon age of child. Cluster 2-6-7-9, Cluster 13-14-18-21, and Cluster 10-23 had relatively large σ^2(sa) components, suggesting that there was no clearly consistent pattern among subjects across the five assessment sessions. Variability due to task difficulty was again negligible.

The addition of items as a facet in the analyses does not, in general, markedly improve the precision of measurement. It seems that a larger portion of observed variance can be attributed to variability due to repeated measures at different ages than can be attributed to items. Table 7 summarizes the results of analyses with increases in number of administrations (N_a), increases in number of scales per cluster (N_i), and simultaneous increases in number of administrations and number of scales. For each cluster, increases in N_a produce a larger generalizability coefficient and a smaller error of measurement [SQRT $\sigma^2(\Delta)$] than do increases in N_i. Furthermore, relative to the effects of increasing N_a, the simultaneous increase of N_a and N_i may not justify the addition of scales to the clusters.

Recommendations

For both the individual scales and the clusters, the question of restricted subject variability needs to be considered. As an estimator of universe score variance, the subject component of variance reflects the ability of the

scales (or clusters) to identify individual differences in teaching interactions. In this regard, it seems advisable to generate additional data with a more heterogeneous sample to resolve the issue concerning the source of low subject variability.

Measurement error in both individual scores and cluster scores can be most effectively reduced by basing evaluation of parent-child interactions during teaching on a series of separate assessments over time. In a practical sense, emphasis should be placed on increasing the number of separate assessments as opposed to concentrating on selecting tasks of different difficulty levels.

References

Barnard, K. Nursing Child Assessment Teaching Scales. Seattle: University of Washington, 1978.

Cronbach, L. J., Gleser, G., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley & Sons, 1972.

Table 1

Teaching Scales and Clusters

Scale Number	Scale Name	Cluster Membership
1	Initial state	
2	Displeasure	
3	Attention	Cluster 3-6-7-9: Readiness to Learn
4	Vocalization	
5	Success	
6	Intensity	Cluster 3-6-7-9: Readiness to Learn
7	Duration	Cluster 3-6-7-9: Readiness to Learn
8	Activity	
9	Alertness	Cluster 3-6-7-9: Readiness to Learn
10	Pos. Feedback	Cluster 10-23: Positive Message
11	Neg. Feedback	Cluster 11-24: Negative Messages
12	Modeling	
13	Materials	Cluster 13-14-18-21: Facilitation
14	Positioning	Cluster 13-14-18-21: Facilitation
15	Directiveness	Cluster 15-16-17: Techniques
16	Forcing	Cluster 15-16-17: Techniques
17	Guidance	Cluster 15-16-17: Techniques
18	Timing	Cluster 13-14-18-21: Facilitation
19	Exploratory	
20	Comfort	
21	Sensitivity	Cluster 13-14-18-21: Facilitation
22	Style	
23	Affection	Cluster 10-23: Positive Messages
24	Disapproval	Cluster 11-24: Negative Messages

Table 2

Individual Scales

Number of Subjects in Each Analysis

Scale	n
1	106
2	106
3	99
4	108
5	107
6	107
7	108
8	107
9	95
10	86
11	82
12	109
13	108
14	110
15	110
16	108
17	109
18	109
19	32*
20	109
21	102
22	105
23	105
24	103

*Scale 19 was not scorable for the majority of subjects.

Table 3

Individual Scales

Variance Components, Generalizability Coefficients,

and Errors of Measurement

Scale	$\sigma^2(sda,e)$	$\sigma^2(da)$	$\sigma^2(sa)$	$\sigma^2(sd)$	$\sigma^2(a)$	$\sigma^2(d)$	$\sigma^2(s)$	$E\rho^2$	SQRT $\sigma^2(\epsilon)$	SQRT $E\sigma^2(\delta)$	SQRT $\sigma^2($
1	.177	.000	.099	.000	.019	.002	.017	.311	.108	.193	.206
2	.504	.010	.176	.000	.091	.040	.036	.297	.159	.292	.354
3	.675	.086	.279	.002	.137	.124	.071	.362	.212	.353	.471
4	.263	.001	.273	.000	.178	.002	.020	.201	.127	.284	.343
5	.844	.062	.065	.000	.000	.173	.022	.186	.135	.312	.436
6	.577	.048	.204	.015	.332	.007	.046	.302	.179	.325	.425
7	.837	.077	.229	.000	.134	.183	.080	.380	.222	.360	.505
8	.313	.009	.284	.000	.041	.017	.008	.082	.085	.297	.325
9	.311	.022	.288	.001	.153	.000	.051	.400	.176	.278	.290
10	.488	.000	.295	.003	.078	.048	.129	.542	.244	.331	.386
11	.363	.119	.132	.000	.000	.013	.060	.489	.175	.250	.285
12	.554	.124	.107	.014	.500	.025	.000	.000	.000	.290	.457
13	.344	.007	.151	.016	.004	.002	.036	.335	.156	.269	.273
14	.295	.005	.147	.002	.011	.005	.037	.380	.151	.244	.254
15	.457	.033	.216	.000	.000	.006	.050	.361	.179	.298	.309
16	.362	.184	.072	.000	.000	.014	.063	.556	.168	.225	.276
17	.384	.261	.194	.001	.000	.000	.060	.408	.189	.296	.337
18	.406	.003	.242	.011	.030	.035	.061	.390	.192	.308	.348
19	.470	.000	.302	.000	.164	.010	.075	.410	.210	.328	.381
20	.351	.006	.400	.013	.194	.003	.034	.217	.163	.349	.403
21	.337	.018	.238	.000	.005	.005	.108	.571	.215	.285	.294
22	.281	.009	.259	.000	.086	.000	.081	.503	.201	.283	.313
23	.351	.003	.259	.039	.085	.038	.060	.362	.196	.326	.377
24	.200	.004	.174	.000	.020	.008	.022	.283	.124	.234	.251

Table 4

Individual Scales

Generalizability Coefficients and Errors

of Measurement When $N_a = 10$

Scale	$E\rho^2$	SQRT $\sigma^2(\varepsilon)$	SQRT $E\sigma^2(\delta)$	SQRT $\sigma^2(\Delta)$
1	.475	.094	.137	.147
2	.458	.140	.207	.238
3	.530	.182	.250	.378
4	.334	.116	.201	.244
5	.313	.124	.221	.372
6	.447	.159	.238	.309
7	.551	.189	.255	.416
8	.152	.082	.210	.238
9	.570	.149	.197	.205
10	.700	.197	.236	.296
11	.657	.143	.177	.210
12	.000	.000	.214	.338
13	.476	.138	.201	.204
14	.547	.129	.174	.184
15	.531	.154	.211	.222
16	.715	.134	.159	.204
17	.578	.159	.210	.239
18	.547	.166	.224	.269
19	.581	.177	.232	.274
20	.344	.149	.254	.292
21	.727	.172	.202	.211
22	.670	.164	.200	.221
23	.490	.175	.250	.300
24	.441	.110	.166	.183

Table 5

Clusters

Number of Subjects and Number of Conditions of Facet i in Each Analysis

Cluster	n	N_i
3-6-7-9	86	4
13-14-18-21	100	4
15-16-17	107	3
10-23	83	2
11-24	77	2

Table 6

Clusters

Variance Components, Generalizability Coefficients,

and Errors of Measurement

	Cluster 3-6-7-9	Cluster 13-14-18-21	Cluster 15-16-17	Cluster 10-23	Cluster 11-24
$\sigma^2(\underline{sida},\underline{e})$	.294	.224	.359	.280	.214
$\sigma^2(\underline{sda})$	.292	.119	.075	.115	.069
$\sigma^2(\underline{ida})$	.014	.003	.049	.003	.036
$\sigma^2(\underline{sia})$	.059	.084	.131	.163	.101
$\sigma^2(\underline{sid})$	.000	.001	.000	.011	.000
$\sigma^2(\underline{da})$	.044	.015	.110	.000	.026
$\sigma^2(\underline{ia})$	.018	.004	.037	.074	.030
$\sigma^2(\underline{id})$	.031	.003	.000	.000	.000
$\sigma^2(\underline{sa})$	.176	.109	.032	.128	.062
$\sigma^2(\underline{sd})$	.009	.005	.004	.007	.000
$\sigma^2(\underline{si})$	.017	.008	.026	.016	.009
$\sigma^2(\underline{a})$	.150	.007	.000	.002	.000
$\sigma^2(\underline{d})$	.042	.005	.000	.034	.014
$\sigma^2(\underline{i})$	.003	.006	1.015	.000	.009
$\sigma^2(\underline{s})$	.043	.051	.031	.091	.033
$E\rho^2$	.339	.514	.409	.557	.424
$\underline{SQRT}(\sigma^2(\varepsilon))$	.168	.157	.136	.276	.137
$\underline{SQRT}(E\,\sigma^2(\delta))$	.289	.219	.212	.269	.211
$\underline{SQRT}(\sigma^2(\Delta))$	.381	.236	.631	.327	.251

Table 7

Clusters

Generalizability Coefficients and Errors of

Measurement with Increases

in N_a and N_i

Cluster	Increases N_a & N_i	$E\rho^2$	SQRT $\sigma^2(\epsilon)$	SQRT $E\sigma^2(\delta)$	SQRT $\sigma^2(\Delta)$
3-6-7-9	when N_a = 10	.481	.149	.215	.300
	when N_i = 8	.359	.165	.276	.367
	when N_a = 10 and N_i = 8	.508	.145	.204	.287
13-14-18-21	when N_a = 10	.659	.132	.162	.179
	when N_i = 8	.547	.152	.205	.220
	when N_a = 10 and N_i = 8	.690	.125	.151	.166
15-16-17	when N_a = 10	.398	.096	.153	.443
	when N_i = 8	.376	.098	.160	.351
	when N_a = 10 and N_i = 8	.491	.089	.127	.325
10-23	when N_a = 10	.654	.178	.220	.263
	when N_i = 4	.598	.191	.247	.287
	when N_a = 10 and N_i = 4	.722	.159	.187	.232
11-24	when N_a = 10	.572	.118	.156	.199
	when N_i = 4	.506	.127	.178	.214
	when N_a = 10 and N_i = 4	.585	.116	.152	.169

SECOND COPY

Technical Report: Generalizability Study of the Nursing Child Assessment Feeding Scales

Daniel W. King and Lynda A. King

Nursing Child Assessment Follow-up Project

Maternal and Child Nursing

University of Washington

Seattle, Washington

November 1978

PS 010927

Daniel King is currently Assistant Professor, Department of Psychology, Central Michigan University, Mt. Pleasant, Michigan.

Technical Report: Generalizability Study of the Nursing Child Assessment Feeding Scales

The Nursing Child Assessment Feeding Scales (NCAFS) consist of 21 separate 7-point Likert-type scales designed to evaluate the quality of parent-child interactions during a feeding situation. Scores for the individual scales are the result of behavioral ratings of both parent and child made by a trained observer. In addition to individual scale scores, scores are obtainable on three clusters or composites of particular scales. Table 1 presents a list of the 21 scales and indicates which scales contribute to specific clusters. A complete description of the scales and the history of their development may be found elsewhere (Barnard, 1978). For the purpose of this report, the scales will be identified by their scale number as shown on Table 1, and the clusters will be identified by their component scale numbers.

For the individual scales, scoring reflects the full 7-point range of the scales; however, for cluster scores, the component scales are "folded" at the mid-point value of 4 (most adaptive behavior) before a composite score is calculated. Thus a value of 5 becomes 3, a value of 6 becomes 2, and a value of 7 becomes 1. With this transformation, the cluster scores then reflect an adaptive-maladaptive continuum.

The present report uses a generalizability approach (Cronbach, Gleser, Nanda, & Rajaratnum, 1972) to the estimation of the precision or reliability of the NCAFS. Unlike the more traditional procedures which examine different types of reliability (test-retest, alternate forms) as independent estimates of precision, the generalizability model allows one to simultaneously evaluate

several factors which are seen as the essential contributors to observed score variance. These factors or "facets" are placed in an analysis of variance design; the resulting data give information on the contributions of specific sources of variance and allow those interested in using the instrument to develop designs which capitalize on the instrument's strengths. Furthermore, a multifaceted generalizability study allows one to examine interactions between facets and thus gain a better understanding of the measurement characteristics of the instrument.

Data for this generalizability analysis are taken from the results of a longitudinal study conducted at the University of Washington designed to identify and develop nursing assessment procedures for significant child development variables. The participants in the study were parents and children who were members of a health-care cooperative in Seattle. Generalizability data are from measures taken at four time points during the child's first year: at ages 1 month, 4 months, 8 months, and 12 months. The NCAFS were one of a series of instruments administered to the participants during assessment sessions conducted in the home.

Individual Scales

The design of the generalizability study of individual scales was a two-facet nested design which may be symbolized as $s \times a(t)$: s represents subjects; t represents a facet identified as food type (milk, solid); and a represents a facet identified as age of child (1 month, 4 months, 8 months, 12 months). At 1 and 4 months of age, children received milk during the feeding session, and at 8 and 12 months of age, they received solid food. Subjects (s) and age of child (a) were treated as random, and food type (t) was considered a fixed facet. Because age of child at administration is

nested within food type, and variance component for age of child [$\sigma^2(\underline{a})$] and the component for the interaction between food type and age [$\sigma^2(\underline{ta})$] are confounded and are thus represented by $\sigma^2(\underline{a},\underline{ta})$. Likewise, the variance component $\sigma^2(\underline{sta},\underline{e})$ is confounded with $\sigma^2(\underline{sa})$. Consideration of $\underline{t}$ as a fixed facet results in the inclusion of any variance attributable to an interaction between subjects and food type as part of the estimate of universe score variance [$\sigma^2(\underline{s})$].

Complete data on all subjects at each administration were required for each analysis, and subjects with incomplete data on any given analysis were eliminated from that analysis. Table 2 provides the number of subjects involved in deriving components of variance for each scale.

Variance components, generalizability coefficients, and errors of measurement are provided in Table 3. In general, variability due to subjects [$\sigma^2(\underline{s})$, the estimate of universe score variance] is low. This finding suggests that the range of the scales used to describe the interaction during the feeding situation is restricted. The most desirable values for $\sigma^2(\underline{s})$ would be in the vicinity of 2.000 to 2.500. Whether restricted range is indicative of sample homogeneity or lack of instrument sensitivity might be clarified in future work with other samples. $\underline{\text{SQRT}}\ \sigma^2(\Delta)$, the appropriate estimate of measurement error for a clinical assessment, must be interpreted with consideration for the size of $\sigma^2(\underline{s})$.

Generally, the component representing the interaction of subjects and food type [$\sigma^2(\underline{st})$] is not a major contributor to observed score variance, suggesting that the pattern of behaviors of subjects is similar across food types. Furthermore, with some exceptions (most noticeably Scales 5, 14, 15, 16, and 17), variability attributed to food type ($\underline{t}$) is a relatively

insignificant part of measurement error. Thus, one might consider elimination of t as a facet in future studies.

The residual component [σ^2(sta,sa,e)] is consistently the largest component of variance and generally much larger than the other two components which comprise observed score variance [σ^2(s) and σ^2(st)]. Among the components of error SQRT $\sigma^2(\Delta)$ [σ^2(t), σ^2(st), σ^2(a,at), and σ^2(sta,sa,e)], it is likewise the largest. Therefore, probably the most efficient method of decreasing error and increasing precision is to increase the number of observations upon which the universe score is based, that is, the number of measures over the child's first year. Table 4 presents errors of measurement and generalizability coefficients when the number of observations per food type (N_a) is increased from two to four. As expected, precision is improved and error is reduced with the exception of those scales where the estimate of universe score equals zero.

Clusters

The clusters were analyzed using a two-facet fully-crossed design, s x a x i, where s is subjects, a is age of the child when the observations were made (1 month, 4 months), and i (items) represents a facet with the conditions being the scales which comprise a cluster. The limited numbers of subjects with complete data (four observations on all items in each cluster) necessitated the elimination of the 8 and 12 month administrations. Table 5 indicates the number of subjects and items for each analysis. Subjects (s), age of child (a), and items (i) were considered random.

The results are contained in Table 6. Subject variability [σ^2(s)] for Cluster 10-11 equals zero, within rounding error; thus, this cluster is eliminated from any further consideration. Subject variability for the other

two clusters suggests that 95% of the ratings of parent-child interactions fell within a range of approximately 1.5 of the 4 possible points on the folded scales.

The residual component [σ^2(si,a,e)] is the largest variance component for Cluster 3-4-5-6-7 and Cluster 1-2. For Cluster 3-4-5-6-7, the data indicate that an increase in precision could best be achieved by an increase in the number of items since component σ^2(si) is larger than component σ^2(sa), both being sources of error in observed score variance. For Cluster 1-2, reliability in the estimation of universe score could be improved by increasing the number of instrument administrations, that is, the number of conditions of the age of child (a) facet.

Recommendations

Future research with the NCAFS should be conducted with samples which demonstrate the potential for resolving the problem of restricted subject variability. The use of samples believed to be more heterogeneous than that one from which the present data were derived should allow the scales an opportunity to identify individual differences in parent-child interactions.

With regard to the individual scales, estimates of the universe score should be based on as many separate observations across time as can be practically incorporated into a study. Whether the observed interaction involves the feeding of milk or solid food appears to be a relatively minor concern. Finally, recommendations for the reduction of measurement error for cluster scores are dependent on which cluster is being considered.

References

Barnard, K. Nursing Child Assessment Feeding Scales. Seattle: University of Washington, 1978.

Cronbach, L. J., Gleser, G., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: John Wiley & Sons, 1972.

Table 1

Feeding Scales and Clusters

Scale Number	Scale Name	Cluster Membership
1	Positioning	Cluster 1-2: Fostering of Growth
2	Focus	Cluster 1-2: Fostering of Growth
3	Kinesthetic	Cluster 3-4-5-6-7: Sensitivity to Cues
4	Visual (Parent)	Cluster 3-4-5-6-7: Sensitivity to Cues
5	Tactile	Cluster 3-4-5-6-7: Sensitivity to Cues
6	Verbal I	Cluster 3-4-5-6-7: Sensitivity to Cues
7	Verbal II	Cluster 3-4-5-6-7: Sensitivity to Cues
8	Mood (Parent)	
9	Tension	
10	Response I	Cluster 10-11: Alleviation of Distress
11	Response II	Cluster 10-11: Alleviation of Distress
12	Initial State	
13	Predominant State	
14	State Changes	
15	Motor Activity	
16	Attentiveness	
17	Vocalization	
18	Mood (Child)	
19	Tone	
20	Visual (Child)	
21	Control	

Table 2

Individual Scales

Number of Subjects in Each Analysis

Scale	n
1	109
2	109
3	151
4	106
5	107
6	109
7	105
8	109
9	108
10	82
11	63
12	107
13	105
14	112
15	108
16	109
17	109
18	109
19	109
20	105
21	103

Table 3

Individual Scales

Variance Components, Generalizability Coefficients,

and Errors of Measurement

Scale	$\sigma^2(\underline{sta},\underline{sa},\underline{e})$	$\sigma^2(\underline{a},\underline{ta})$	$\sigma^2(\underline{st})$	$\sigma^2(\underline{t})$	$\sigma^2(\underline{s})$	$E\rho^2$	$\frac{SQRT}{\sigma^2(\varepsilon)}$	$\frac{SQRT}{\sigma^2(\delta)}$	$\frac{SQRT}{\sigma^2(\Delta)}$
1	.317	.020	.006	.054	.037	.313	.160	.286	.336
2	.729	.025	.041	.037	.050	.197	.200	.451	.477
3	.726	.052			.292	.445	.402	.603	.624
4	.579	.001	.024	.002	.032	.167	.162	.396	.397
5	.844	.130	.045	.193	.110	.320	.273	.484	.602
6	.509	.000	.157	.042	.442	.682	.375	.454	.477
7	.689	.015	.089	.000	.316	.593	.359	.466	.470
8	.544	.000	.000	.000	.000	.000	.000	.369	.369
9	.466	.000	.000	.010	.098	.457	.231	.341	.349
10	1.247	.000			.000	.000	.000	.790	.790
11	.880	.000	.000	.023	.000	.000	.000	.269	.481
12	.681	.000	.044	.091	.027	.123	.154	.439	.488
13	.521	.126	.025	.034	.020	.121	.132	.378	.438
14	2.790	.021	.001	.282	.022	.031	.146	.836	.919
15	.582	.192	.081	.155	.026	.124	.152	.431	.558
16	.650	.091	.000	.238	.035	.175	.169	.403	.552
17	.960	.079	.000	.479	.044	.155	.193	.490	.707
18	.726	.015	.016	.101	.002	.009	.042	.435	.494
19	.226	.038	.019	.000	.022	.247	.128	.257	.218
20	.540	.017	.233	.025	.069	.217	.233	.501	.518
21	.808	.005	.000	.000	.087	.302	.247	.449	.451

Table 4

Individual Scales

Generalizability Coefficients and Errors

of Measurement When $N_a = 4$

Scale	$E\rho^2$	SQRT $\sigma^2\{\varepsilon\}$	SQRT $E\sigma^2\{\delta\}$	SQRT $\sigma^2\{\Delta\}$
1	.468	.141	.206	.268
2	.308	.186	.335	.365
3	.616	.334	.426	.441
4	.272	.151	.291	.292
5	.462	.243	.358	.491
6	.757	.328	.377	.404
7	.707	.304	.362	.364
8	.000	.000	.261	.261
9	.628	.191	.241	.252
10	.000	.000	.558	.558
11	.000	.000	.332	.349
12	.201	.147	.327	.391
13	.202	.125	.279	.333
14	.060	.144	.591	.702
15	.189	.146	.337	.463
16	.298	.156	.285	.460
17	.268	.179	.346	.608
18	.018	.042	.314	.389
19	.364	.117	.194	.206
20	.274	.224	.429	.445
21	.463	.216	.318	.319

Table 5

Clusters

Number of Subjects and Number of Conditions of Facet i in Each Analysis

Cluster	n	N_i
3-4-5-6-7	138	5
1-2	153	2
10-11	65	2

Table 6

Clusters

Variance Components, Generalizability Coefficients, and Errors of Measurement

Cluster	σ^2(sia,e)	σ^2(ia)	σ^2(sa)	σ^2(si)	σ^2(a)	σ^2(i)	σ^2(s)	$\underline{E}\rho^2$	SQRT $\sigma^2(\epsilon)$	SQRT $\underline{E}\rho^2(\delta)$	SQRT $\sigma^2(\Delta)$
3-4-5-6-7	.611	.005	.114	.162	.005	.063	.181	.546	.527	.388	.407
1-2	.517	.019	.078	.038	.001	.000	.072	.279	.228	.433	.439
10-11	.730	.000	.580	.066	.000	.000	.000	.000	.000	.711	.711